

Die zehn Open-Data-Kriterien der *Sunlight-Foundation*¹

1) Vollständigkeit

Von Regierung und Verwaltung veröffentlichte Datensätze sollten so vollständig wie möglich sein und den ganzen Umfang dessen abbilden, was zu einem bestimmten Thema dokumentiert ist. Sämtliche Rohdaten eines Datensatzes sollten der Öffentlichkeit zur Verfügung gestellt werden, mit Ausnahme dessen, was Gesetze zum Schutz personenbezogener Daten gebieten. Metadaten, die die Rohdaten beschreiben und erklären, sollten zusammen mit Formeln und Erklärungen zur Berechnung der Daten ebenfalls mitgeliefert werden. Dies wird den Benutzern erlauben, die Ausrichtung der verfügbaren Information zu verstehen und jedes Datum mit dem größtmöglichen Detailreichtum zu untersuchen.

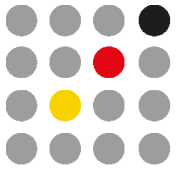
2) Primärquellen

Von Regierung und Verwaltung veröffentlichte Datensätze sollten Primärquellen sein. Dies schließt die ursprünglich von der Regierung erhobenen Informationen ein, sowie Details darüber, wie die Daten gesammelt wurden und die ursprünglichen Quelldokumente, die die Erhebung dokumentieren. Die öffentliche Verbreitung ermöglicht es den Benutzern zu überprüfen, dass die Informationen korrekt erhoben und genau aufgezeichnet wurden.

3) Zeitliche Nähe

Von Regierung und Verwaltung veröffentlichte Datensätze sollten der Öffentlichkeit innerhalb eines angemessenen Zeitraums zur Verfügung stehen. Wenn machbar, sollten von der Regierung erhobene Informationen veröffentlicht werden, sobald sie erhoben und zusammengestellt sind. Vorrang sollten solche Daten erhalten, deren Nützlichkeit zeitabhängig ist. Echtzeit-Updates würden den Nutzen maximieren, den die Öffentlichkeit aus diesen Informationen ziehen kann.

¹ Englisch Original: "Ten Principles for Opening Up Government Information" (2007): <http://sunlightfoundation.com/policy/documents/ten-open-data-principles/>, Deutsche Übersetzung leicht abgewandelt nach: http://wiki.opendata-network.org/Ten_Principles_for_Opening_Up_Government_Information



4) Leichter Zugang

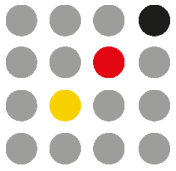
Von Regierung und Verwaltung veröffentlichte Datensätze sollten so zugänglich wie möglich sein. Zugänglichkeit ist dabei als die Einfachheit, mit der Informationen eingeholt werden können – sei es auf dem elektronischen oder auf dem physischen Weg – definiert. Hürden zum physischen Zugang schließen die Notwendigkeit ein, persönlich ein bestimmtes Büro aufsuchen oder bestimmte Abläufe erfüllen zu müssen (wie das Ausfüllen von Formularen oder das Einreichen von IFG-Anfragen). Hürden zum automatisierten elektronischen Zugang beinhalten den Zugang zu Daten nur über ausgefüllte Eingabemasken oder Systeme, die browserorientierte Technologien erfordern (z.B. *Flash*, *Javascript*, *Cookies* oder *Java Applets*). Im Gegensatz dazu macht eine Benutzerschnittstelle, die das Herunterladen sämtlicher gespeicherter Daten auf einmal (*bulk access*) erlaubt, oder Angebote, bestimmte Daten über eine Programmierschnittstelle (*API*) abzurufen, die Daten viel zugänglicher. (Ein weiterer Aspekt ist die Auffindbarkeit, also die Möglichkeit, einfach Inhalte zu finden und herunterzuladen.).

5) Maschinenlesbarkeit

Maschinen können mit bestimmten Arten von Eingaben viel besser umgehen als mit anderen. Zum Beispiel sind handschriftliche Notizen auf Papier sehr schwer für Maschinen zu verarbeiten. Einscannen von Text über Texterkennungsprogramme (*Optical Character Recognition, OCR*) produziert viele Zuordnungs- und Formatierungsfehler. Auch Informationen, die im bekannten PDF-Format verbreitet werden, sind für Maschinen sehr schwer analysierbar. Deswegen sollten Informationen in etablierten Dateiformaten abgespeichert werden, die maschinenlesbar sind. Wenn andere Faktoren den Einsatz schwer maschinenlesbarer Formate erfordern, sollten die Daten zusätzlich in maschinenfreundlichen Formaten verfügbar sein. Dateien sollten von einer Dokumentation begleitet werden, die sich auf das Format bezieht und darauf, wie man es in Bezug auf die Daten verwendet.

6) Diskriminierungsfreiheit

Diskriminierungsfreiheit bezieht sich darauf, wer auf Daten zugreifen kann und wie dieser Zugriff erfolgt. Hürden bei der Datennutzung umfassen Registrierung oder Mitgliedschaftsvoraussetzungen. Eine weitere Hürde ist die Verwendung von „umzäunten Gärten“, wenn also nur bestimmte Anwendungen auf die Daten zugreifen dürfen. Im weitesten Sinn bedeutet diskriminierungsfreier Zugang, dass jede Person zu jeder Zeit auf die Daten zugreifen kann, ohne sich identifizieren zu müssen oder eine Rechtfertigung für ihr Handeln abgeben zu müssen.



7) Verwendung offener Standards

Die Forderung nach der Nutzung gemeinsam entwickelter („offener“) Standards bezieht sich auf das Eigentum an den verwendeten Formaten. Wenn zum Beispiel nur eine einzige Firma ein Programm vertreibt, mit dem eine Datei in einem bestimmten Format gelesen werden kann, ist der Zugang zu diesen Informationen abhängig von der Nutzung der Software dieser Firma. Manche dieser Programme sind für die Allgemeinheit nicht zugänglich oder nur gegen Gebühr erhältlich. *Microsoft Excel* zum Beispiel ist ein weit verbreitetes Tabellenkalkulationsprogramm, dessen Nutzung Geld kostet. Häufig existieren kostenlos verfügbare Formate, durch die auf die Daten zugegriffen werden kann, ohne eine Software-Lizenz zu benötigen. Wenn diese Kosten-Hürden beseitigt werden, sind die Daten für eine größere Gruppe potenzieller Nutzer verfügbar.

8) Lizenzierung

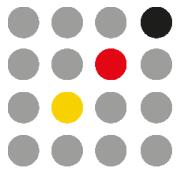
Das Auferlegen von Nutzungsbedingungen, zum Beispiel die Pflicht zur Namensnennung, Einschränkungen bezüglich der Verbreitung usw. wirken als Hürden für die Verwendung von Daten. Maximale Offenheit bedeutet, öffentliche Daten – ggf. als Werk der Regierung oder Verwaltung ausgewiesen – ohne Nutzungsbeschränkungen gemeinfrei verfügbar zu machen.

9) Dauerhaftigkeit

Die Möglichkeit, Informationen über lange Zeit hinweg zu finden, wird als Dauerhaftigkeit bezeichnet. Von Regierung und Verwaltung online veröffentlichte Daten sollten beständig (*sticky*) sein, da heißt, sie sollten in Archiven dauerhaft online verfügbar sein. Häufig werden Informationen aktualisiert, verändert oder entfernt, ohne einen Hinweis darauf zu geben, dass sich etwas geändert hat. Oder Informationen werden als Datenstrom verfügbar gemacht, aber nicht archiviert. Um eine Verwendung der Daten durch die Öffentlichkeit zu befördern, sollten einmal online gestellte Informationen mit angemessener Versionskontrolle und dauerhafter Archivierung online bleiben.

10) Nutzungskosten

Eine der größten Hürden beim Zugriff auf scheinbar öffentlich verfügbare Informationen sind die Kosten, die der Öffentlichkeit für den Zugriff auferlegt werden – auch wenn diese minimal sind. Regierung und Verwaltung haben zahlreiche Grundlagen, um der Öffentlichkeit den Zugang zu ihren eigenen Dokumenten in Rechnung zu stellen: Die Kosten für das Erzeugen der Informationen; eine Kostendeckungs-Grundlage (die Kosten für das Produzieren der Informationen geteilt durch die erwartete Zahl der Käufer); die Kosten für das Abrufen der Informationen; eine Pauschale pro Seite oder Anfrage; Bearbeitungsgebühren; die Kosten für die Vervielfältigung etc.



Die meisten Regierungsinformationen werden für Regierungszwecke erhoben und die Nutzungskosten haben geringe bis gar keine Auswirkungen darauf, ob Regierung und Verwaltung die Daten erheben. Das Berechnen von Gebühren für den Zugang verzerrt die Menge derer, die gewillt (oder fähig) sind, auf Informationen zuzugreifen. Nutzungskosten können zudem die Weiterverarbeitung von Daten, die Wirtschaftswachstum und Steuereinnahmen generieren, von vornherein ausschließen.

STAND Mai 2014

VERSION 1.0

LIZENZ Diese Informationen sind vom Bundesministerium des Innern erstellt worden und stehen unter der Lizenz „Creative Commons Namensnennung 3.0 Deutschland“.

